AD_____

Award Number: W81XWH-07-1-0044

TITLE: Identification of the Transformational Properties and Transcriptional Targets of the Oncogenic SRY Transcription Factor SOX4

PRINCIPAL INVESTIGATOR: Christopher Scharer

CONTRACTING ORGANIZATION: Emory University
Atlanta, GA 30322

REPORT DATE: January 2008

TYPE OF REPORT: Annual Summary

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
10-01-2008	Annual Summary	11 DEC 2006 - 10 DEC 2007

4. TITLE AND SUBTITLE
Identification of the Transformational Properties and Transcriptional Targets of the Oncogenic SRY Transcription Factor SOX4

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-07-1-0044

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Christopher Scharer

Email: cdschar@learnlink.emory.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Emory University
Atlanta, GA 30322

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
The exact role of SOX4 in development and promoting tumorigenesis however is currently unknown. Here we sought to identify the direct transcriptional targets of SOX4 on a global scale to determine the gene networks affected in human cancers and development. Using chromatin immunoprecipitation coupled to DNA microarrays tiling the promoters of 25,000 known genes (ChIP-chip), we identified 140 high confidence promoter regions bound by SOX4 in living human prostate cancer cells. We have also used a unique protein-binding double-stranded DNA microarray to determine a novel SOX4 specific position-weight matrix for *in silico* SOX4 binding site searches. Direct targets of SOX4 include several key cellular regulators and 11 other transcription factors such as SOX11, ZNF281, and ZHX2. SOX4 impacts the Notch pathway, FGF signaling via regulation of FGFRL1, as well as the Hedgehog pathway via regulation GLIS2. These data provide new insights into how SOX4 impacts growth factor and developmental signaling pathways and how these changes may influence cancer progression and development.

15. SUBJECT TERMS
ChIP-chip, SOX4, Protein-binding microarray,

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
U	U	U	UU	14	19b. TELEPHONE NUMBER *(include area code)*

Table of Contents

Introduction:

SOX4 is a critical and required regulator of development and recent evidence implicates a role for SOX4 in carcinogenesis. The goal of this research is to evaluate the transcriptional and oncogenic properties of the transcription factor SOX4 and to determine its role in murine prostate development. Our lab has previously shown SOX4 mRNA and protein to be overexpressed in prostate cancer, and this expression is correlated with increasing Gleason score. Other labs have shown SOX4 to be overexpressed in other tumors such as leukemia, melanoma, glioblastoma and bladder carcinomas. Despite this knowledge little is known of the direct transcriptional targets of SOX4, and how misregulation of these networks affects human cancers and development. To determine the direct transcriptional targets on a global scale we performed chromatin immunoprecipitation coupled to DNA microarrays. We used human promoter arrays from Nimblegen, Inc. that tiled roughly 5 kb of promoter and intronic sequence for 25,000 known genes. Total coverage for the array was roughly 110 Mb of DNA. Using this technique we were able to determine the direct SOX4 targets in living prostate cancer cells. We have also obtained a SOX4 floxed mouse that will enable the prostate specific deletion of SOX4 in mice. This information will determine if SOX4 is required for the development of a functional prostate. Determining the transcriptional targets and *in vivo* functions of SOX4 will contribute critical knowledge to the SOX4 field.

Body:

AIM 1: Determine the Direct Transcriptional Targets of SOX4 on a Global Scale using a ChIP-chip approach.

In order to facilitate chromatin immunoprecipitation (ChIP) of SOX4, an HA epitope tag was inserted onto the N-terminus and the HA-SOX4 construct was cloned into an eYFP expressing lentiviral vector (Figure 1A). RWPE-1 and LNCaP prostate cancer cell lines were stably infected with the lentivirus and fluorescence-activated cell sorting (FACS) analysis was used to purify and pure population of eYFP expressing cells (Figure 1B). Two cell lines were created expressing either an eYFP only control construct or the HA-SOX4 and eYFP genes. Immunoprecipitation (IP) was performed to ensure HA-SOX4 was expressed and could be IPed using our 12CA5, anti-HA monoclonal antibody (Figure 1C). Both RWPE-1 and LNCaP cell lines were created but only the LNCaP-HA-SOX4 and control cells were used for the ChIP-chip experiment. ChIP assays were performed in triplicate for the LNCaP-HA-SOX4 cell lines and in duplicate using the control cell lines. DNA was extracted and purified using standard ChIP protocols from Nimblegen, Inc. IPed and total Input DNA was amplified using the ligation-mediated PCR approach and 4 ug of total DNA was sent to Nimblegen, Inc for the labeling and hybridization reactions.

Signal intensities were z-score normalized, log2 transformed and ratios of IPed to total Input signal calculated for each probe set. To identify enriched peaks ChIPoTle analysis (*1*) was carried out using a window of 500 bp and a step size of 50 bp. ChIPoTle software uses a sliding window approach to look for peaks that are enriched across multiple neighboring probes and assigns a p-value for a genomic region based on a Gaussian error function. Peaks that overlapped in two of the three data sets, were not present in the LNCaP-YFP cell line and scored a p-value less than $1x10^{-5}$ were called significant (Figure 2A). Using this approach 139 genes contained significant overlapping peaks and were labeled direct SOX4 targets (Table 1). To verify the set of 139 direct SOX4 target genes, 10 candidate SOX4 target genes were chosen at random, QRT-PCR primers were designed around the peaks and enrichment was verified by conventional ChIP (Figure 2B). All 10 of the genes were reproducibly enriched in the LNCaP-HA-SOX4 cell line as well as the RWPE-1 cell line over the YFP control (Figure 2B). We further validated 6 more genes that met our p-value criteria in both the LNCaP and RWPE-1cell lines by PCR (Figure 2C and 2D). All genes tested were enriched in both cell lines except ANKRD15, which was not enriched in the RWPE-1 cell line. These results confirm the validity of our data set.

HMG domain transcription factors bind AT rich DNA in the minor groove and two previous reports identify a 7mer SOX4 binding motif (*2, 3*). While this knowledge can aid in the search for putative binding sites it does not take into account the role of alternate bases at various positions. A SOX4 specific position-weight matrix is required to fully utilize the power of bioinformatic searches. Apart from the consensus core SOX family binding site WWCAAW, where W represents either A or T, little is known about what preferences SOX4 exhibits at each base position during binding (*4*). In order to facilitate bioinformatic searches for SOX4 DNA binding sites we sought to determine a SOX4 specific position-weight matrix (PWM) using a unique, protein-binding, double stranded DNA microarray (*5*). The array allows recombinant protein to interact with and bind every possible 10mer, thus allowing *in vitro* binding site specificities to be calculated. We generated an N—terminal, GST-SOX4-DBD fusion protein, and expressed and purified it from *E. coli* (Figure 3B). To ensure the purified recombinant fusion protein was functional we performed an electromobility shift assay (EMSA) using a published SOX4 binding site of AACAAAG (*2*). Increasing concentrations of GST-SOX4-DBD was incubated with radiolabeled specific probe alone, with a cold specific competitor or a cold non-specific competitor. GST-SOX4-DBD was able to bind the probe and cause a shift that was abolished when cold specific competitor probe, but not when cold non-specific probe was added (Figure 3A). These data show that the truncated GST-SOX4-DBD fusion protein is functionally active *in vitro*. The GST-SOX4-DBD was incubated with the protein binding microarray and a novel PWM ($AACAA^{A}/_{T}\ ^{G}/_{A}\ ^{G}/_{A}/_{C}$) was calculated according to published protocols (Figure 3C) (*5*). Two groups have previously reported similar binding site sequences for SOX4: AACAAAG (*2*) and AACAAT (*3*). Our PWM confirms both of the previous known

binding sites and adds new information on the binding preferences in the 8th position as well as alternate bases at the 6th and 7th positions.

Using our newly synthesized PWM, we applied CONFAC (*6*) software to analyze the enriched sequences for the presence of SOX4 binding sites. We analyzed the sequences for the enriched peaks in the promoters of our 139 verified genes as well as 18 YFP enriched control sets containing peaks of equal sequence length. With stringent criteria (core similarity ≥ 0.85, matrix similarity ≥ 0.75) we find 83 of 139 (60%) contain at least one SOX4 binding site, and those peaks that contain SOX4 binding sites have on average 4 SOX4 sites per peak. SOX4 binding sites were significantly enriched relative to 18 sets of random, YFP enriched sequences ($p < 0.0019$ by Mann-Whitney U-test and Benjamini correction for multiple hypothesis testing). Previous studies have also implicated that SOX proteins mediate their transcriptional activity by interacting with other transcription factors such as the SOX2-OCT3/4 pair (*4*). CONFAC software was used again to search the sequences for the presence of co-occurring transcription factor binding sites. Using the same criteria as before, comparing the verified sequences to 18 random controls we determined that the E2F family was the most frequently co-occurring site with a q-value of 1.91×10^{-8} (Table 2). Interestingly, GO ontology analysis of the 139 SOX4 target genes revealed that 6% of them are involved in cell cycle. This finding suggests that part of SOX4's function is to control the expression of cell cycle-regulated genes (Figure 4A). Other co-occurring transcription factor binding sites that were overrepresented are the WHN and HEB, a forkhead and TCF transcription factor respectively. SOX4 has been previously shown to modulate WNT signaling via interaction with β-catenin and a TCF transcription factor, suggesting a possible role for SOX4 in transcriptionally modulating WNT signals (*7*).

In order to determine the biological processes and functions of the SOX4 targets we performed a GO ontology analysis using GOstat software (*8*). GOstat analysis annotates gene lists with GO functions and calculates an enrichment p-value with corrections for multiple hypothesis testing. GOstat analysis identified several highly enriched biological functions, including anatomical development, transcriptional regulation, protein folding, signal transduction, cell cycle regulation, angiogenesis, and cell motility. A similar gene ontology analysis using DAVID software (*9*) of the list of direct SOX4 targets found that the top annotated biological process was transcription ($p = .024$) and the top annotated molecular functions were nucleic acid ($p = 6.6x^{10-4}$) and DNA binding ($p = 2.9x^{10-3}$). Interestingly, the analysis identified 11 other transcription factors (Table 3) as SOX4 regulatory targets suggesting that SOX4 may regulate other transcriptional networks. Ingenuity Pathway Assist (IPA) analysis identified biological pathways and functions that are enriched in our verified gene list compared to random control lists. IPA analyses discovered key components of the EGFR, Notch, AKT-PI3K and WNT-Bcatenin pathways as SOX4 regulatory targets. Using this information we built a SOX4 regulatory network found in prostate cancer cells (Figure 4B). SOX4 target genes comprise key components such as ligands (DLL1 and NGR1), a regulatory kinase (PDPK1) and downstream transcription factors (FOXO3 and HES2). These data suggest that SOX4 impacts key developmental and growth factor signaling pathways in prostate cancer cells.

AIM2: Determine the effects of Loss or Overexpression in vivo

Our collaborator, Dr. Neal Copeland at the National Cancer Institute (NCI), generated a mouse containing a LOX-STOP-LOX-SOX4 allele inserted into the Rosa26 genomic locus (Figure 5A). This construct, when crossed to a mouse expressing Cre recombinase under the control of a prostate specific probasin promoter, causes the STOP codon to be excised allowing expression of the SOX4 transgene (Figure 5B). This allows the prostate specific overexpression of SOX4. Under the supervision of the Emory animal facility Rosa26-SOX4 mice were breed to Probasin-Cre mice to generate a line of Rosa26-SOX4/Probasin-Cre mice. The construct predicts that GFP expression should be lost and SOX4 expression gained. However, upon analysis of prostate RNA by quantitative real-time PCR (QRT-PCR) we detected a decrease in GFP mRNA but surprisingly no significant increase in SOX4 mRNA levels (Figure 5C and 5D). Pathological evaluation of prostate sections revealed no structural abnormalities and immunohistochemical staining for the presence of

SOX4 protein did not show a difference between controls and Rosa26-SOX4/Probasin-Cre mice (Figure 6). Either through biological selection for lower SOX4 levels or a technical problem with expression from the Rosa26 locus, SOX4 overexpression was not detected in these mice. Unfortunately this project was discontinued and will not be studied further. Future directions include using a Tet-inducible SOX4 that does not require genomic rearrangements may provide further insights into the effects of SOX4 overexpression.

The other side to this study is to specifically knockout SOX4 in the prostate to determine the developmental requirement for SOX4. We recently received these mice from Dr. Neal Copeland. The endogenous SOX4 allele is flanked by LOXP sites. When crossed to the Probasin-Cre mice, endogenous SOX4 will be excised allowing us to study the effects of loss of SOX4 on the prostate. The Probasin promoter becomes active around birth, continues into adulthood and spans the majority of prostate development (*10*).

Key Research Accomplishments:

- Determined 139 high confidence direct SOX4 target genes
- Identified a novel PWM for SOX4
- Incorporated the PWM into bioinformatic analysis to find SOX4 binding sites
- Identified the possible pathways that SOX4 influences
- Evaluated the Rosa26-SOX4 mice for SOX4 overexpression and the phenotypic consequences
- Established the initial breeding for the prostate specific SOX4 knockout mice

Reportable Outcomes:

- *Manuscripts:* The research performed as described in AIM 1 is currently being prepared for submission to *Genome Research* and will be submitted in early 2008.
- *Abstracts:* The research performed as described in AIM1 will be presented as a poster for the 2008 Keystone Meeting – Signaling Pathways in Cancer and Development
- *Presentations:* All research preformed during the training grant will be presented annually at an internal department seminar as part of my graduate training program.

Conclusion:

The SOX4 field has become interesting in the last couple of years due to the recent evidence linking SOX4 to multiple developmental processes and cancers. However, despite being a transcription factor there is little knowledge about the direct SOX4 target genes and the transcriptional networks SOX4 affects. This information is critical to understanding the downstream effects of SOX4. My research has vastly expanded on previous knowledge of SOX4 transcriptional targets; identifying 139 high-confidence genes. Future work will be required to verify how SOX4 affects the predicted pathways and what the phenotypic consequences are of having too much SOX4, in the case of various cancers, or no SOX4 at all which has shown severe developmental consequences in mice.

In vivo work attempting to overexpress SOX4 alone has not shown any obvious complications to prostate development, although we were never able to show that SOX4 is in fact overexpressed in our model system. To this end a tetracycline inducible system may be developed in the future which will allow us to control SOX4 levels and will allow us to study SOX4 overexpression with a different system. The future work involving the SOX4 knockout mice will be extremely interesting given that all other tissues where SOX4 has been knocked out have shown severe consequences.

References:

1. M. J. Buck, A. B. Nobel, J. D. Lieb, *Genome Biol* **6**, R97 (2005).
2. M. van Beest *et al.*, *J Biol Chem* **275**, 27266 (Sep 1, 2000).
3. M. van de Wetering, M. Oosterwegel, K. van Norren, H. Clevers, *Embo J* **12**, 3847 (Oct, 1993).
4. V. Lefebvre, B. Dumitriu, A. Penzo-Mendez, Y. Han, B. Pallavi, *Int J Biochem Cell Biol* **39**, 2195 (2007).
5. M. F. Berger *et al.*, *Nat Biotechnol* **24**, 1429 (Nov, 2006).
6. S. Karanam, C. S. Moreno, *Nucleic Acids Res* **32**, W475 (2004).
7. D. Sinner *et al.*, *Mol Cell Biol* **27**, 7802 (Nov, 2007).
8. T. Beissbarth, T. Speed, *Bioinformatics* **20**, 1464 (Jun 12, 2004).
9. G. Dennis, Jr. *et al.*, *Genome Biol* **4**, P3 (2003).
10. X. Wu *et al.*, *Mech Dev* **101**, 61 (Mar, 2001).

Publications and Abstracts

2006. P. Liu, S. Ramachandran,, M. Ali-Seyed, **C.D. Scharer**, N. Laycock, W. B. Dalton, H. Williams, S. Karanam, M. W. Datta, D. L. Jaye, and C. S. Moreno**.** SOX4 is a Transforming Oncogene in Human Prostate Cancer Cells. Cancer Research, 66**:** 4011-4018.

Ali-Seyed, M, **C.D. Scharer**, and C.S. Moreno. SOX4 Participates in an Epidermal Growth Factor Receptor Positive Feedback Loop [Abstract]. Mechanisms & Models of Cancer, Cold Spring Harbor Laboratory, August 16-20, 2006

2008. **C.D. Scharer**, J.F. McDonald, B.B. Benigno, and C.S. Moreno. Inhibition of Aurora Kinases Overcomes Resistance to Paclitaxel in Ovarian Cancer. In Preparation

2008. **C.D. Scharer**, C.D. McCabe, M. Ali-Seyed, M.F. Berger, M.L. Bulyk, and C.S. Moreno. Genome-wide Location Analysis of the SOX4 Transcriptional Network in Prostate Cancer. Genome Research. In Preparation

Supporting Data:

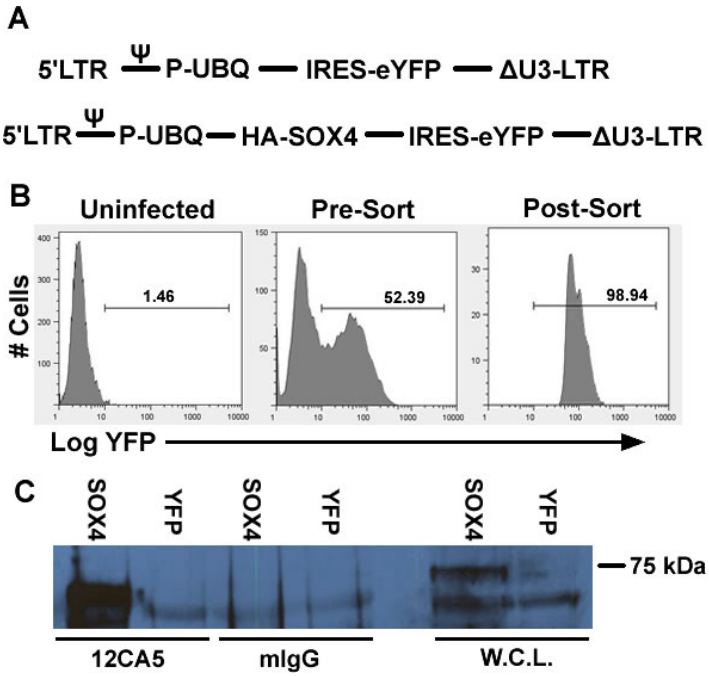

Figure 1: **(A)** Schematic diagram of the lentiviral constructs used to stably infect LNCaP and RWPE-1 prostate cancer cells showing the locations of LTRs and promoters. The top figure represents the control, eYFP only construct, and the lower figure represents the HA-SOX4 construct. **(B)** Histogram charts showing the control uninfected, pre-sorted and post-sorted cell populations. Lower axis displays YFP signal intensity. **(C)** Immunoblot showing that HA-SOX4 is expressed and specifically immunoprecipitated from the LNCaP-HA-SOX4 cell line and not the control LNCaP-YFP cell line.

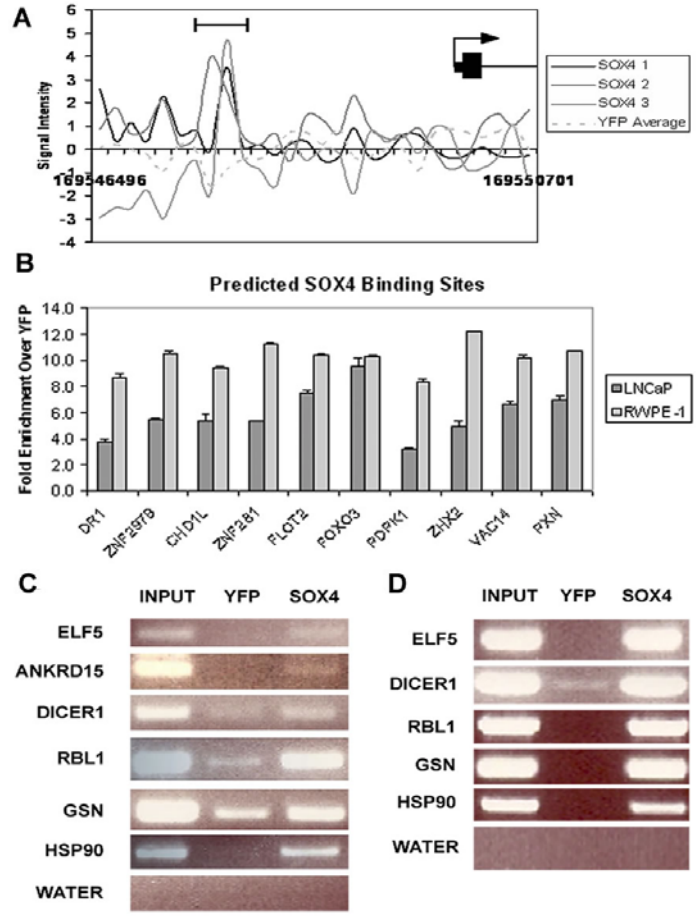

Figure 2: **(A)** Graph showing enrichment in the three HA-SOX4 lanes over the average of the two YFP replicates for the gene FMO4. **(B)** QRT-PCR analysis of 10 randomly selected genes verified in both the RWPE-1 and LNCaP cell lines. Graph shows fold enrichment of the HA-SOX4 IP over the YFP control IP. **(C)** Genes that were verified by conventional ChIP assay. LNCaP-HA-SOX4 and LNCaP-YFP cells were subjected to conventional ChIP followed by PCR in both the LNCaP and RWPE-1 prostate cell lines.

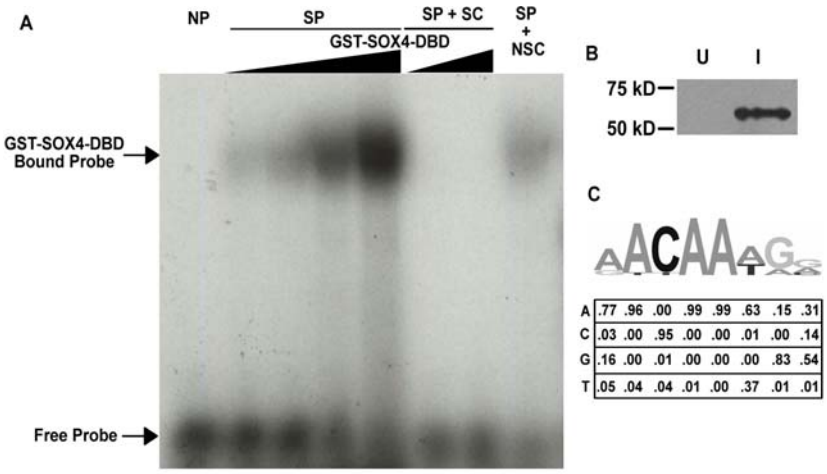

Figure 3: **(A)** EMSA assay of recombinant GST-SOX4-DBD binding to a known SOX4 binding motif of a 35mer oligo. NP – No protein, SP – specific probe, SC – Specific cold competitor, NSC – non-specific cold competitor. **(B)** SDS-PAGE gel of GST-SOX4-DBD from an IPTG uninduced (U) or induced (I) cell line. **(C)** Novel 8mer PWM for SOX4 displayed both graphically and numerically for each base position

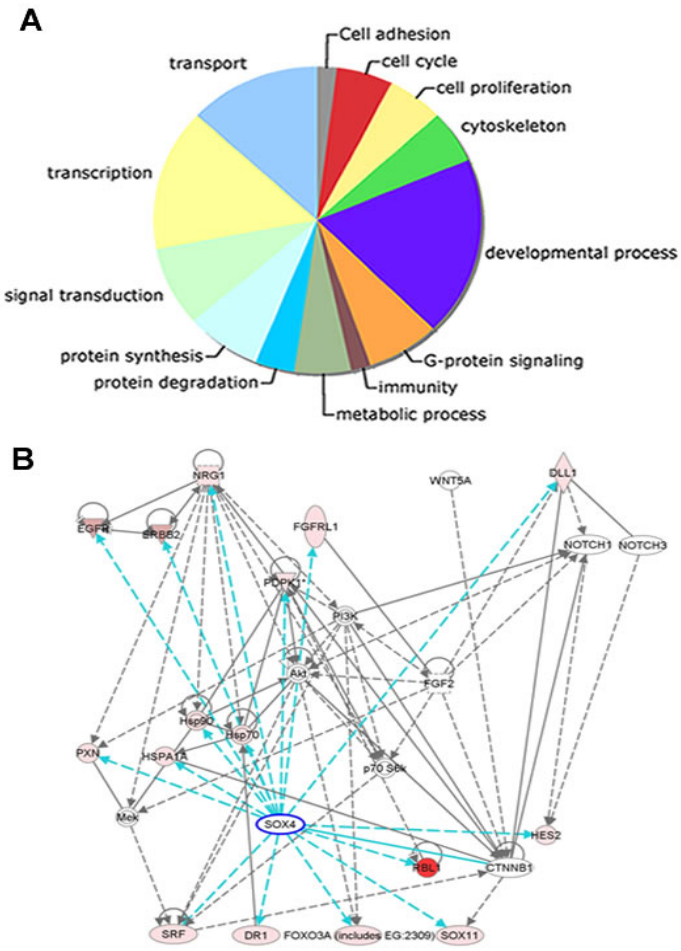

Figure 4: **(A)** Pie chart generated from GOstat analysis showing the biological function of SOX4 target genes. **(B)** Ingenuity Pathway Assist analysis showing SOX4's transcriptional network.

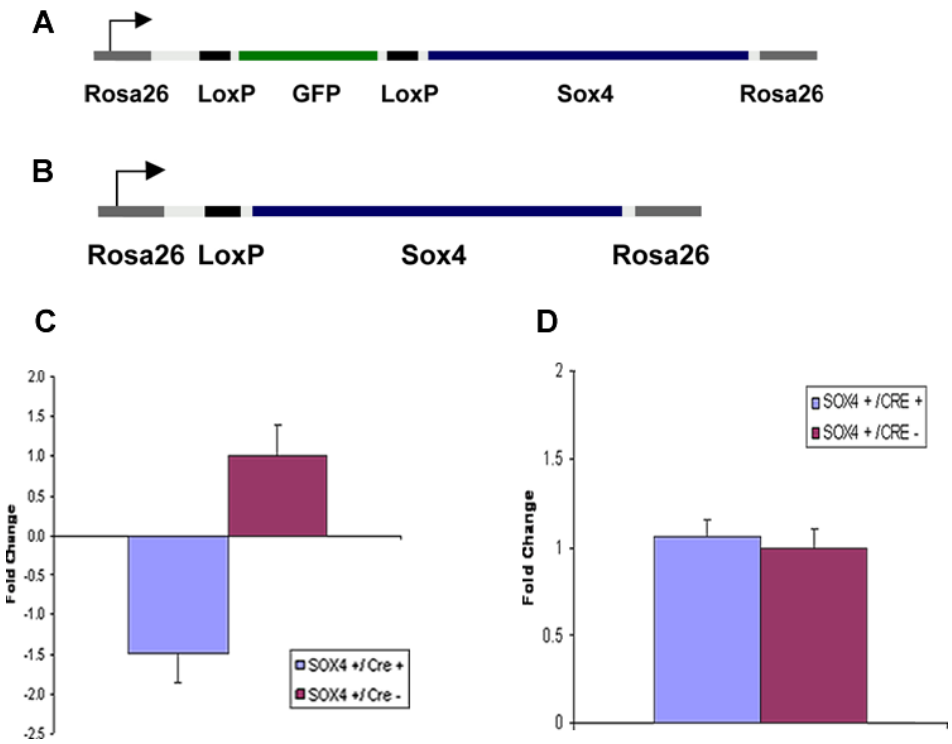

Figure 5: **(A)** Schematic showing the genomic arrangement of the Rosa26 locus with the SOX4 construct inserted. **(B)** Schematic showing the Rosa26 locus after Cre recombination allowing SOX4 expression. **(C)** QRT-PCR analysis of prostate RNA for the presence of GFP mRNA. As expected GFP expression is lost when Cre is expressed. (N=8) **(D)** QRT-PCR analysis of SOX4 expression in mouse prostates. SOX4 is not overexpressed as expected in the Cre positive mice. (N=8)

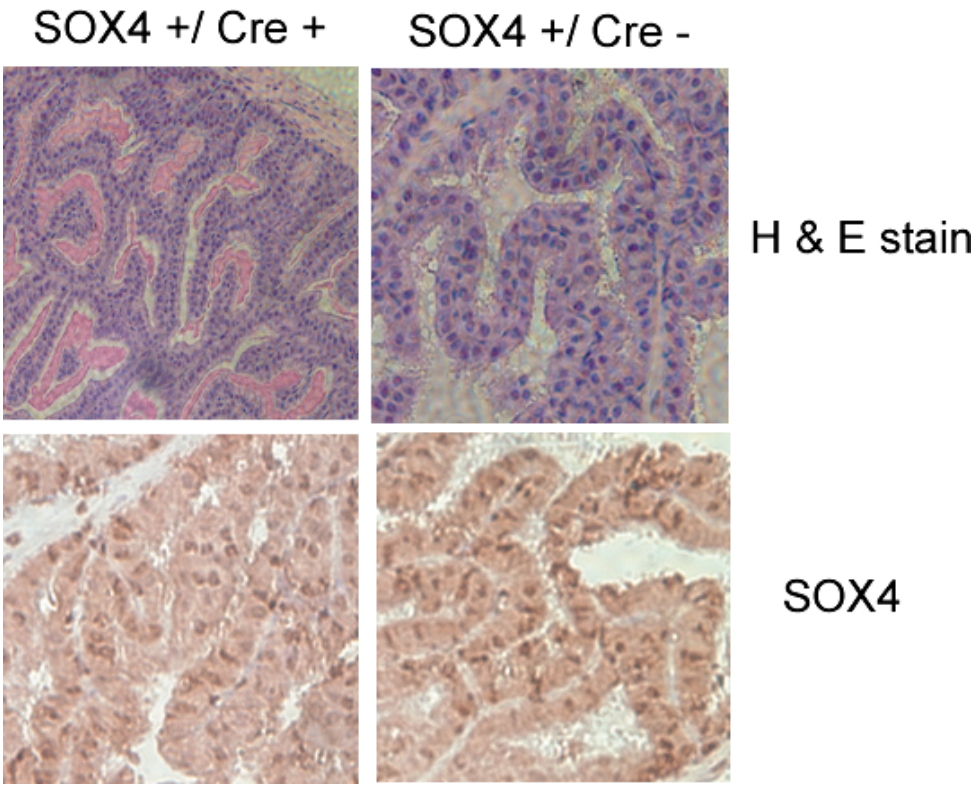

Figure 6: Cross sections of mouse prostates from SOX4 +/Cre + and SOX4 +/Cre – mice. Top panel shows H & E staining while bottom panel is stained with a monoclonal antibody to SOX4. There are no morphological differences or differences in SOX4 levels between the two samples.

Table 1: 139 verified SOX4 target genes.

Symbol	Entrez ID	p-value	Symbol	Entrez ID	p-value	Symbol	Entrez ID	p-value
ACVR2A	92	1.10E-10	HES2	54626	3.81E-09	WDR20	91833	2.00E-09
ADCY5	111	7.63E-07	HLA-DRA	3122	2.26E-09	WDR26	80232	1.40E-07
AFF4	27125	7.62E-07	HMGA2	8091	3.84E-07	WDR51A	25886	1.48E-07
AGTPBP1	23287	4.57E-08	HSP90AA1	3320	2.00E-09	ZBTB43	23099	9.95E-05
AHCYL1	10768	4.58E-06	HSPA1A	3303	3.00E-11	ZDHHC21	340481	3.90E-05
ALDH18A1	5832	8.24E-10	HSPA1L	3305	3.00E-11	ZHX2	22882	1.91E-12
ANAPC13	25847	3.67E-10	KCND2	3751	6.40E-07	ZMYND10	51364	1.14E-10
ANKRD15	23189	2.60E-09	KIAA0329	9895	8.23E-09	ZNF271	10778	3.62E-11
ANKRD34	284615	2.10E-07	KIAA1033	23325	1.10E-12	ZNF281	23528	4.08E-08
ARHGAP24	83478	1.85E-06	KIAA1804	84451	2.64E-07	ZNF509	166793	2.68E-06
ARPC5L	81873	4.81E-08	LANCL2	55915	2.49E-10	ZRANB3	84083	1.11E-07
C11orf56	84067	1.91E-15	LDLRAP1	26119	8.88E-07			
C17orf42	79736	1.29E-13	LHFPL2	10184	5.16E-09			
C1orf121	51029	4.49E-05	LHPP	64077	6.59E-08			
C1orf14	81626	2.24E-08	LIX1L	128077	2.10E-07			
C20orf112	140688	1.14E-10	LOC124216	124216	1.15E-08			
C5orf21	83989	1.80E-13	LOC126075	126075	3.37E-10			
C6orf89	221477	2.07E-08	LOC158301	158301	5.16E-05			
C9orf102	56959	1.48E-06	LOC284513	284513	2.02E-06			
CAMSAP1L1	23271	4.46E-05	LOC414300	414300	1.10E-12			
CDH24	64403	6.99E-07	LPPR2	64748	3.37E-10			
CEP63	80254	3.67E-10	LYAR	55646	2.68E-06			
CGGBP1	8545	2.85E-07	MARCH5	54708	4.31E-10			
CHD1L	9557	1.46E-08	METTL5	29081	7.53E-05			
CHIC2	26511	5.70E-06	MGC3205	90585	3.37E-10			
CINP	51550	8.23E-09	MYF5	4617	2.88E-14			
CLGN	1047	5.25E-06	NRP1	8829	8.63E-05			
CNGA4	1262	1.91E-15	OGG1	4968	1.50E-07			
COG2	22796	6.84E-07	OR11H12	440153	1.50E-06			
COMMD8	54951	8.54E-06	OR8K1	390157	1.77E-09			
CORO2A	7464	6.43E-07	OXGR1	27199	4.27E-07			
CPEB3	22849	4.31E-10	PDPK1	5170	1.15E-08			
DICER1	23405	2.42E-05	PEX16	9409	5.63E-07			
DLL1	28514	1.87E-11	PMP22CD	338661	2.61E-07			
DMKN	93099	7.44E-09	PMS2L3	5387	1.84E-06			
DR1	1810	8.09E-07	POLR3GL	84265	2.10E-07			
DRD3	1814	1.41E-06	POU5F2	134187	2.20E-06			
DSG4	147409	2.18E-09	PPP2R5C	5527	1.70E-06			
EDG3	1903	2.75E-05	PRDM16	63976	6.39E-06			
EEF1D	1936	9.86E-07	PRSS3	5646	1.57E-08			
EIF1B	10289	8.61E-07	PXN	5829	9.97E-10			
ELF5	2001	4.05E-06	R3HDM1	23518	1.11E-07			
EPN3	55040	3.85E-06	RAB3D	9545	3.37E-10			
ESPN	83715	3.81E-09	RAET1G	353091	1.72E-13			
ESPNL	339768	8.48E-09	RASSF1	11186	4.47E-12			
EVI5L	115704	9.21E-21	RBL1	5933	3.74E-21			
FAM3B	54097	1.57E-07	RBM34	23029	4.82E-08			
FBXO4	26272	1.99E-08	RGS20	8601	4.66E-07			
FGFRL1	53834	8.46E-06	RHOQ	23433	1.30E-06			
FHAD1	114827	1.92E-07	RNF19	25897	9.51E-06			
FLJ33065	440952	8.61E-07	RPL35	11224	4.81E-08			
FLJ41757	440862	1.30E-06	SCGB2A2	4250	7.77E-09			
FLJ42875	440556	6.39E-06	SLC44A1	23446	7.16E-08			
FLOT2	2319	2.89E-13	SOX11	6664	1.06E-06			
FMO4	2329	1.79E-09	STRBP	55342	4.08E-06			
FOXO3	2309	1.13E-06	TAAR9	134860	2.59E-09			
GALNT14	79623	2.16E-07	TBC1D2B	23102	5.12E-06			
GLIS2	84662	8.75E-08	TEKT2	27285	6.31E-06			
GNA14	9630	1.63E-10	TIGD5	84948	9.86E-07			
GPR110	266977	4.80E-06	TMEM57	55219	6.02E-06			
GSN	2934	5.31E-11	TUSC2	11334	4.47E-12			
GSTA3	2940	1.25E-06	UBR4	23352	2.02E-06			
GYLTL1B	120071	5.63E-07	UNQ501	374882	3.37E-10			
HELT	391723	3.03E-08	VAC14	55697	2.18E-14			

Table 2: Benjamini corrected q-values for co-occurring transcription factor binding sites

Transcription Factor	Family	Benjamini Corrected q-value
E2F1	E2F	1.91E-08
MAZ	MAZ	1.92E-08
HEB	TCF	2.74E-08
NFKAPPAB	NF-KB	3.21E-08
WHN	Forkhead	4.87E-08
PAX5	Paired Box	5.38E-08
ELK1	ETS	1.16E-07
SMAD4	SMAD	3.49E-07
CREB	CREB	1.18E-06
CMYB	MYB	2.87E-06

Table 3: Transcription factors regulated by SOX4

Symbol
AFF4
DR1
ELF5
FOXO3
GLIS2
HMGA2
MYF5
PRDM16
SOX11
ZHX2
ZNF281